科技论文引用对象研究综述*

马娜[1,2] 张智雄[1,2,3] 于改红[1,2]
[1] 中国科学院文献情报中心 北京 100190 [2] 中国科学院大学图书情报与档案管理系 北京 100190
[3] 中国科学院武汉文献情报中心 武汉 430071

摘要：［**目的/意义**］为更好地提升基于内容的引文分析效果，对国内外引用对象相关研究进行调研总结，为引用内容分析研究提供借鉴。［**方法/过程**］通过调研国内外引用对象相关研究，梳理引用对象的概念定义、分类体系、应用领域和自动化识别等方面研究进展，总结当前引用对象研究不足并提出未来发展方向。［**结果/结论**］引用对象从语义层面评价文献学术研究的贡献和利用价值，为引文分析方法增加了重要维度。引用对象研究需要从理论、技术和应用三个方向进行深化：理论上，加强多维度引用对象特征的研究和分析；技术上，探索基于大规模语料的自动化识别方法；应用上，尝试基于引用对象的科研评价服务。

关键词：引用对象 引用对象识别 引用内容分析 科研评价
分类号：G250
DOI：10.13266/j.issn.0252－3116.2019.23.016

1 引言

科技论文是科研创新成果的重要载体之一，作者在撰写论文过程中会引用他人的研究成果，这种引用行为反映了科研工作之间复杂的关联关系，是引文分析的基础。传统引文分析方法仅仅考虑引用次数，无法揭示文献间的内容关联和本质联系。

针对这一问题，一些学者开始对引用内容进行分析[1－4]。引用内容分析指对引用行为的语义信息进行挖掘和分析，已有研究主要围绕引用动机识别[5－7]、情感倾向分析[8－10]、引用位置分析[11－12]、引用内容主题分析[13－14]、引文内容抽取[15－17]、引用内容分析框架研究[18－19]等方面展开。这种揭示引用行为语义信息的研究方法弥补了传统引文分析方法以数量衡量质量的不足。有学者认为，基于引用内容的引文分析将成为下一代引文分析的方向[1－2]。

在引文分析向引用内容分析方向深化的同时，引用对象的概念被提了出来。通过调研发现，引用对象有多种表述，如 Concept Symbols（概念符号）、Knowledge Claims（知识声明）、Reference Terms（参考文献术语）等，也有研究者直接定义为“引用对象”[10,19]，从概念定义上看，不同名称下本质并无差异。简而言之，引用对象指的是引文被引用的具体科学内容，它直接揭示了文献间的本质联系。引用对象从语义层面评价文献学术研究的贡献和利用价值，为引文分析方法增加了重要维度。引用对象的定性分析与传统基于被引频次的定量引文分析方法结合，从内容与数量相结合的方式完善科研评价体系，是针对被诟病多年的以计数方式为主的科研评价的一剂良方，具有重要的理论意义和实际应用价值。

鉴于引用对象的重要意义和作用，为了能够更全面地了解引用对象研究相关进展，笔者使用英文“citation、reference”分别与“object、factoid、terms、concept symbols、context”等词语进行组配，中文“引文、引用、对象、内容、参考文献、术语、关键词”等关键词进行组配，在权威学术网站和数据库文献检索平台进行检索，对相关研究文献及研究报告进行深入阅读和关联阅读，总结科技论文引用对象相关研究进展。从理论层面梳理引用对象基本概念，对比不同分类体系差异；从应用层面总结引用对象在引文分析和文本挖掘中的利用；从技术层面介绍引用对象自动识别方法、效果和不足。

* 本文系中国科学院文献情报能力建设专项项目“基于arXiv数据的物理领域科研论文自动语义标注和索引应用示范”（项目编号：院1657）研究成果之一。

作者简介：马娜（ORCID：0000－0001－5016－0879），馆员，硕士；张智雄（ORCID：0000－0003－1596－7487），中国科学院武汉文献情报中心主任，研究员，博士，通讯作者，E-mail：zhangzhx@mail.las.ac.cn；于改红（ORCID：0000－0003－1301－2871），馆员，硕士。

收稿日期：2019－01－22 修回日期：2019－04－23 本文起止页码：139－145 本文责任编辑：杜杏叶

最后综合引用对象研究进展，分析当前研究存在的问题，展望未来引用对象研究的发展方向。

2 引用对象的概念研究

目前研究者对引用对象的定义并不统一，但对其本质认知高度一致，即在引用内容中表达引文内容的对象，常以名词或名词术语的形式表示。不同学者给予的定义略有差别，为了纵观引用对象研究脉络，笔者按时间先后顺序逐一梳理研究中引用对象概念。

引用对象概念雏形最早由 H. Small 在 1978 年提出，H. Small 在论文中详细介绍了引用对象的内涵和表现形式，认为引用内容中的科学内容（scientific content）是作者对被引文献的内容进行符号化的过程，以一种泛化的“概念符号”（concept symbols）描述引文中的概念或方法[20]。他认为这种概念符号比引文分类更加直接地解释了作者的引用行为，具有重要价值。

A. Ritchie 等利用引文中术语进行索引优化时，把引用句中与引文相关的术语称为参考文献术语（reference terms），并对术语的表现形式进行分析[21]。V. Qazvinian 等在论文摘要任务中把引用句里描述引文主要贡献的无重叠的词组称为事实（non-overlapping contribution，fact）[22]。在其随后的论文中，用 nugget 表示[23]，含义未变，fact 和 nugget 即为引用对象。

M. H. Anderson 等基于 H. Small 对引用对象的描述，总结引文原文的知识声明（knowledge claims），并与引用句中描述引文的具体内容进行对照，映射到每个引用句中的知识声明即为引用对象[24]。许德山在其《科技论文引用中的观点倾向分析》博士论文中，把引用句中出现的来自于引文的最大字符串定义为引用对象[10]。R. Jha 等在自动摘要研究时，利用引用句中关于引文的事实（factorid）[25]构建自动摘要模型，虽然在文中未给出 factorid 定义，但笔者观察其数据集发现 factorid 与 V. Qazvinian 等定义的 fact 含义一致，即能够表达引文内容的名词术语。

陆伟等在引文内容标注框架研究中，把引用对象标注体系作为其研究重点之一，定义引用对象为作者在引用文献时提及成果中的具体内容对象[19]。A. Khalid 等对比了引用句中描述引文的术语与采用主题识别算法获得的主题词的一致性，用参考文献术语（reference terms）表示与引文相关的内容即引用对象[26]。

综合上述研究成果，不同研究者对引用对象的理解基本一致，笔者总结引用对象概念，即引用对象是作者在引用文献时提及引文中的具体内容对象，由施引文献发起，是作者根据研究需要对引文内容的一种泛化表示，具有概括性和客观性的特点。引用对象可能是引文中的具体方法、工具、数据、结果等对象（如句 a），也可能是对引文中方法、观点、应用或结论的概括性总结（如句 b）。

（a）<u>*Atomic scattering factors*</u> *for the refinement of both structures were taken from Hanson, et al.*

（b）*Brown et al.* （1993） *proposed a* <u>*series of statistical models of the translation process.*</u>

进一步研究科技论文与其引文语义层面引用关系，笔者认为其遵循以下的基本模型：

引用关系 = {论文，引用句，引文，引用对象，引用频次（一次、多次），引用目的，情感倾向（肯定、中立、否定）}

其中引用句是指论文中描述引文内容的句子，引用目的表示引用的动机，情感倾向包括肯定、否定和中立，引用频次可能是一次或者多次，而引用对象则是论文中描述引文具体内容对象，是论文与引文内容上的关联点，直接表明引文被利用价值和学术贡献。

基于上述引用关系基本模型，可以明确看到，引用对象与引用目的、引用情感倾向等的相关关系，有利于更好地明确引用对象的概念。

3 引用对象的分类研究

引用对象的分类研究主要从两个角度展开，一是对引用对象的表现形式分类，二是对引用对象的内容分类。许德山按不同表现形式把引用对象分为命名实体引用对象和抽象表述引用对象[10]，命名实体对象主要以名词或名词短语等结构体现的算法、术语、概念、数据、软件等，抽象表述对象主要以句子或子句形式表示引文的观点、理论、方法等。命名实体引用对象与 H. Small 研究的“概念符号”的表现形式一致，目前已有研究主要针对命名实体对象展开[20-26]。

从内容层面对引用对象进行分类较为复杂，目前没有形成统一的分类体系，研究者按照其对引用对象内容的理解把引用对象分为 5-11 个类别不等，并通过人工标注实验，分析不同类型引用对象出现的频次及占比，更深入、系统地揭示引用对象特点。

M. Garzone 分类体系定义了 5 种类型引用对象，分别为“材料、设备、工具”“公式”“方法、程序”“条件、限制”“结果分析方法”[27]。R. Radoulov 对 M. Garzone 的体系进行修正[28]，构建的分类体系既包括表示引用目的的对象如“研究背景”“先驱性工作”“相关工作”等，

又包括具体内容对象如“模型”“产品或材料”“数据”等。陆伟等进一步改进了 R. Radoulov 的分类体系[19]，摒弃描述引用目的对象，增加具象类型如“工具”“公式/推导”，细化“方法”类为“方法”“模型”和“算法”，保留了“概念”“数据”“结果”类型，最终形成 11 个类目的引用对象分类体系。

由上可见，分类的标准和细粒化程度对分类体系影响较大。笔者认为对引用对象的分类应体现引用对象内容本质特征，而非其在被引用时的作用或目的，分类标准应具有独立性和唯一性，否则容易造成类目混淆。如引用对象为某个“数据集”，从施引文献引用目的角度可能为了表示“研究背景”或“已有工作”，混合两种分类标准会导致同一引用对象属于多个类型，分类结果一致性较差。除分类标准不同，已有分类体系的细粒度具有较大差异，如同样是“方法”类，陆伟等将其细化为“方法”“模型”和“算法”三个类目，其他研究者则统一定义为方法类，并不做更近一步的区分。细化的分类有利于充分了解引用对象但也容易带来一些问题，如一些概念所属的类型在不同文章中并不相同，如“Expectation Maximization”在不同的文章中被述为“方法”或者“算法”[19]。

4 引用对象的识别研究

引用对象识别是一种基于非结构化数据的信息抽取任务。作者在引用文献时所使用的语言模式较为灵活，在某些情况下，甚至连人工确认引用对象都比较困难。许多学者表示，引用对象自动识别是一个比较困难的任务[22－23,29]。目前引用对象识别可分为两大类，人工标注识别和利用计算机模型进行自动化识别。已有研究以人工识别为主，自动化识别研究较少。

4.1 人工识别

引用对象人工识别指通过人工标注的方式识别引用句中的引用对象。笔者从数据集和标注规则两个角度梳理 H. Small、A. Ritchie、V. Qazvinian、M. H. Anderson、陆伟、R. Jha 等学者的人工标注引用对象方法如表 1 所示：

表 1 人工标注采用数据集及标注规则

作者	领域	数据量	标注规则
R. Small	化学	294 种化学期刊中前 52 篇高被引论文和其全部引用内容	标记与引文最相关术语或短语
A. Ritchie 等	计算机－计算语言学	1 篇论文及其 24 条引用内容	标注引用句中描述引文内容的术语
V. Qazvinian 等	计算机－计算语言学	25 篇 ACL 论文及其 954 条引用内容	阅读每篇论文的全部引用句，抽取论文的主要贡献；归一化不同标注者标注的贡献并与引用句进行对应
M. H. Anderson	管理学	1 篇论文及其 496 条引用内容	阅读论文归纳“知识主张”，结合 496 条引用内容中的表述，总结出 16 个主要引用对象
陆伟等	计算机	20 篇 LDA 模型论文及其引用内容	标注引用内容中描述引文具体内容对象
R. Jha 等	计算机－计算语言学	47 篇论文及其 2625 条引用内容	标注引用句中与引文相关的事实(去掉只出现一次的事实)

(1)数据集方面，由于没有引用对象开放数据集，学者们主要采用自建数据集或利用科技文献全文数据集，数据内容主要包括引文与其引用内容，数据量方面，受限于人工标注成本，标注数据量并不大，引用对象数量从几十到上千不等。受到较多学者青睐的计算机语言协会数据集(Association for Computational LinguisticsAnthology Network corpus，ANN)[29]是引文内容分析常用数据集之一，它提供自然语言处理领域权威期刊和顶级会议学术论文全文，并建立丰富的引用网络。

(2)标注规则方面，基本标注规则一致度较高，即标注引用句中与引文相关内容，但由于后续的应用需求不同，标注策略有所不同。H. Small、A. Ritchie 等、陆伟等、R. Jha 等直接标注每一条引用句中的引用对象；V. Qazvinian 等把所有相同引文的引用句聚集在一起进行统一标注；M. H. Anderson 从所有引用句中总结出所有引用对象再分配给每条引用句。不同的标注策略下，最终标注结果的数据形式有所不同，第一种标注策略下引用对象完全来自引用句且表达形式丰富，后两种策略引用对象表现形式较为规范化，可直接应用于深度引用分析。

4.2 自动化识别

通过计算机技术方法实现引用对象自动化识别正处于起步阶段，R. Radoulov、许德山和 A. Khalid 三位研究者先后从引用对象的分类、识别和替代识别三种不同角度进行探索和尝试，下文逐一梳理这些技术方法和实验效果。

R. Radoulov 采用机器学习算法实现引用对象分

类，算法融合了论文结构特征、引文位置特征、词性特征、句法特征和两类线索词，为每种对象类型训练朴素贝叶斯分类器模型，再用训练好的模型实现引文分类自动标注[28]。数据集包含5 583条引文数据，结果显示，在9种类别中，特定背景、数据、方法和产品4种类别的分类效果较好，准确率70%以上，平均F值在0.7－0.5之间。其他5种类别一般背景、相关工作、概念、历史性说明和开创性工作，分类效果并不理想，F值只有0.3左右，R. Radoulov分析原因认为这种结果是由于数据量不足导致的，这几种类别的数据仅占数据集总量的10%。从这个角度也可以看出，引用对象的类型主要集中在特定背景、数据、方法和产品这4种类别。

许德山分析科技论文引用情感链路时对引用对象进行识别。识别方法为计算引用句字符串与引文原文内容的相似性，抽取引用句与引文共同含有的最大字符串作为引用对象，实验中由于引文全文获取难度较大，匹配工作仅基于引文标题[10]。实验语料由61篇科技文献组成，标注语料去重后得到1 076条引证标记，最终正确识别引用对象705个，正确率为65.52%。除引用对象识别外，许德山尝试对一篇文献的不同引用对象进行规范化处理，通过对引用对象词形相似度、字符串长度和语义相似度加权求和计算同一篇引文的两个引用对象是否为同一对象，引用对象规范化处理程度直接影响引用对象应用效果。

A. Khalid通过对一篇引文的所有引用内容集合的主题进行识别，替代难度较大的直接识别引用对象工作。他从ANN论文集中选取113篇论文和9 970条引用句，邀请一名计算语言学研究者标注引用句与引文相关的术语，如果存在二义性则放弃标注[26]。随后，在每篇论文的引用句集合上，采用LDA主题模型进行主题识别，利用词云为每个主题分配一个可以概括主要内容的高频术语词作为主题标签，并对相同内容的主题进行合并。最后评估主题标签与人工标注术语间的相似性，结果表明在大多数情况下，引用句集合的主题词可以表示引用句中的参考文献术语，即引用对象。但也有例外，如引用句中包含多于一个引用对象的情况。

基于引用对象识别研究现状的梳理，该研究处于起步阶段，没有一套完整的研究体系。在数据集方面，研究者各自创建实验数据，引用对象自动识别目前并没有公开测评预料，不利于对比不同方法下的识别效果。在技术方法上，只有许德山真正尝试在引用内容中抽取引用对象，但仅考虑字符串匹配的方法过于简单，识别效果并不理想。A. Khalid探索利用LDA模型生成主题词代替直接识别引用对象，但该模型忽略了词与词之间的顺序，没有考虑词汇是否与引文真正相关等问题。引用对象自动化识别需要更多的探索和尝试。

5 引用对象在引文分析和文本挖掘中的应用

引用对象表示他人对引文研究成果的引用，具有知识表示和知识传播两种特征。引用对象的利用主要分为两种情况，一种是利用引用对象与被引频次相结合进行引文分析，从知识传播和继承的角度，分析引文的学术贡献和影响。另一种是利用引用对象作为重要特征，融入文本挖掘算法，提高文献检索、自动摘要等应用效果。目前引用对象的应用主要基于人工标注数据，受限于标注成本，应用规模较小，仅处于探索和尝试阶段。

科技文献对学术界的主要贡献可以通过学术共同体对其引用和引用它时的描述信息揭示，引用对象天生承载着这部分信息，在引文分析和科研评价方面具有先天优势。M. H. Anderson等利用引用对象和被引频次从定性和定量相结合的方式，分析一篇管理学高被引论文的学术贡献[24]。引用对象的标注过程，首先汇总高被引论文的全部引用句，总结被提及的主要贡献，再把这些贡献映射回每条引用句，作为该引用句的引用对象。这样的标注方法在标注的同时规范了引用对象的表达，有利于后期深入利用分析。

应用方面，M. H. Anderson等把标注的引用对象与出现频次相结合分析了高被引论文的主要学术贡献，并且通过分别计算七个不同学科的施引文献中引用对象占比情况，比较发现该高被引论文在不同学科的影响并无较大差异。尤其在管理和计算机这两个主要学科，三个最常被引用的引用对象非常相似。这种内容与数量联合测度论文贡献的方法弥补了传统引文分析方法的不足，为科学计量及科研评价提供了新思路。

除增强引文分析方法外，引用对象体现了其他研究人员对引文成果的总结和刻画，利用这一特点引用对象被应用在文献检索、自动摘要、情感分析等领域。A. Ritchie等认为引用句中与引文相关的术语在构建一篇引文的索引词时具有重要作用，尝试把引用文献的引用对象与该文献本身的索引词结合起来，并举例证明引用对象的加入更好的描述了论文内容[21]。V.

Qazvinian 等在利用引用内容自动生成论文摘要时，利用引用对象确定引文的重要贡献，除去重复引用内容，提高自动摘要的准确性和完整性[22]；R. Jha 等在 V. Qazvinian 的启发下，利用引用对象及其出现的次数作为权重加入论文摘要生成模型，有效地提升了摘要生成效果[25]。许德山在研究语篇链路引用情感时，首先识别引用对象并进行对象消歧，通过信息在句子间传递的方式构建引用对象语篇链路进而完成引文评价情感链路分析[10]。

6 总结与展望

引用对象从内容层面揭示引文的学术价值，弥补了传统以数量评价质量的引文分析方法的不足，具有重要的理论意义和应用价值。本文从理论、应用和技术三个层面梳理科技论文引用对象研究进展，不难看出引用对象的概念在逐步明确，分类体系研究初见成效，但在自动化识别和应用方面存在明显不足。总体来讲，引用对象研究正处于初期探索阶段，已有研究远不能实现引用对象大规模应用的目标，还有许多研究值得进一步探索与深化。

在理论上，从多维度加强引用对象特征的研究和分析，进一步明确引用对象的本质。一个科学问题在研究初期，理论研究待逐步完善，引用对象需要更多理论研究支撑大规模识别和应用。引用对象作为知识泛化表示对象，在表现形式上具有统计学特征，包括词性特征、在引用句中承担的句法成分、与引用标签的位置关系、语言模型特点等。从语言学角度深入分析引用对象特点，有助于深入、客观地理解引用对象。另外，不同学科论文的引用对象特点分析也是一个有价值的研究问题。受全文获取限制，已有研究大部分基于计算机领域，引用对象在其他学科特征是否一致不得而知，多领域的引用对象特征对比研究非常必要。

在技术上，探索基于大规模语料的引用对象自动化识别方法。目前引用对象自动化识别成果十分有限，笔者认为主要有两方面原因，一方面由于引用对象是施引文献作者对引文内容的转述，主观性较强；另一方面缺少大规模高质量的测评语料，不利于研究的开展。技术方面，笔者认为引用对象识别可以转化为序列标注问题，可以尝试 HMM[30]、CRF[31]、LSTM-CNNs-CRF[32]、BERT[33]等在序列标注问题上取得不错效果的模型。数据集方面，基于科技文献构建具有一定规模的高质量研究数据集，由论文、引文、引用句和引用对象组成的结构化数据，具有数据结构清晰、样本形式丰富、标注质量高、易于获取等特点。针对目前训练数据量较少的情况，可以考虑采用半监督学习[34]或迁移学习[35]等不同策略，适当降低模型对训练数据量的要求。

在应用上，借助自动化识别技术，提升语义化引文分析应用效果，尝试基于引用对象的科研评价服务。引用对象体现了论文在科学研究中的学术价值和贡献，这是其他引用内容不具备的。引用对象定性分析与传统基于被引频次的定量引文分析方法结合，完善科研评价体系，是引用对象最根本也是最重要的应用方向。自动化识别效果的不断提升，为深入内容的引文分析奠定坚实基础，使基于引用对象的科研评价服务成为可能，同时从应用角度反观技术需要改进之处，二者相互促进。利用大规模引用对象建立语义化实体级引用网络，可以快速实现学术论文贡献分析，挖掘领域经典理论、重要算法或广泛应用的数据集，探索交叉学科学术交叉点，构建知识流动路径。

参考文献：

[1] DING Y, ZHANG G, CHAMBERS T, et al. Content-based citation analysis: the next generation of citation analysis[J]. Journal of the Association for Information Science and Technology, 2014, 65(9):1820－1833.
[2] 赵蓉英，曾宪琴，陈必坤. 全文本引文分析——引文分析的新发展[J]. 图书情报工作, 2014, 58(9): 129－135.
[3] 刘浏，王东波. 引用内容分析研究综述[J]. 情报学报, 2017, 36(6):637－643.
[4] 胡志刚. 全文引文分析[M]. 北京：科学出版社,2016.
[5] GARFIELD E. Can citation indexing be automated? [J]. National Bureau of Standards,1964,269(114):189－192.
[6] TEUFEL S, SIDDHARTHAN A, TIDHAR D. Automatic classification of citation function[C]//Proceedings of the 2006 conference on empirical methods in natural language processing. Sydney: Association for Computational Linguistics, 2006:103－110.
[7] MUNKHDALAI T, LALOR J, YU H. Citation analysis with neural attention models[C]// Proceedings of the seventh international workshop on health text mining and information analysis (LOUHI). Austin: Association for Computational Linguistics, 2016:69－77.
[8] ATHAR A, TEUFEL S. Context-enhanced citation sentiment detection[C]// Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: human language technologies. Montreal: Association for Computational Linguistics, 2012:597－601.
[9] YU B, FANELLI D. Classifying negative findings in biomedical publications[J]. Association for Computational Linguistics, 2014: 19－23.

[10] 许德山. 科技论文引用中的观点倾向分析[D]. 北京:中国科学院文献情报中心,2012.
[11] DING Y, LIUX, GUO C, et al. The distribution of references across texts: some implications for citation analysis[J]. Journal of informetrics, 2013, 7(3): 583 – 592.
[12] 胡志刚, 陈超美, 刘则渊, 等. 从基于引文到基于引用——一种统计引文总被引次数的新方法[J]. 图书情报工作, 2013, 57(21): 5 – 10.
[13] LIU S, CHEN C. The differences between latent topics in abstracts and citation contexts of citing paper[J]. Journal of American Society for Information Science and Technology, 2013, 64(3): 627 – 639.
[14] 祝清松, 冷伏海. 基于引文内容分析的高被引论文主题识别研究[J]. 中国图书馆学报, 2014, 40(1): 39 – 49.
[15] ANGROSH M A, CRANEFIELD S, STANGER N. Conditional random field based sentence context identification: enhancing citationservices for the research community[C]//Proceedings of thefirst Australasian Web conference-volume 144. Adelaide: Australian computer society, 2013: 59 – 68.
[16] SONDHI P, ZHAI C X. A constrained hidden markovmodelapproach for non-explicit citation context extraction[C]//Proceedingsof the 2014 Society for Industrial and Applied Mathematics international conference on data mining. Pennsylvania: Society for industrial and applied mathematics, 2014: 361 – 369.
[17] 雷声伟, 陈海华, 黄永, 等. 学术文献引文上下文自动识别研究[J]. 图书情报工作, 2016, 60(17): 78 – 87.
[18] ZHANG G, DING Y, MILOJEVIC S. Citation content analysis (cca): A framework for syntactic and semantic analysis of citation content[J]. Journal of the American Society for Information Science and Technology, 2013, 64(7): 1490 – 1503.
[19] 陆伟, 孟睿, 刘兴帮. 面向引用关系的引文内容标注框架研究[J]. 中国图书馆学报, 2014, 40(6): 93 – 104.
[20] SMALL H. Cited documents as concept symbols[J]. Social studies of science, 1978, 8(3): 327 – 340.
[21] RITCHIE A, TEUFEL S, ROBERTSON S. How to find better index terms through citations [C]//Proceedings of COLING/ACLworkshop on how can computational linguisticsimproveinformation retrieval. Sydney: Association for Computational Linguistics, 2006: 25 – 32.
[22] QAZVINIAN V, RADEV D R. Scientific paper summarization using citation summary networks[C]//Proceedings of the 22nd international conference on computational linguistics. Manchester: Association for Computational Linguistics, 2008: 689 – 696.
[23] QAZVINIAN V, RADEV D R, OZGUR A. Citation summarization through keyphraseextraction[C]//Proceedings of the 23rd international conference on computational linguistics. Beijing: Association for Computational Linguistics, 2010: 895 – 903.
[24] ANDERSON M H, SUN Y T. What have scholars retrieved from Walsh and Ungson (1991)? a citation context study[J]. Management learning, 2010, 41(2): 131 – 45.
[25] JHA R, ABU-JBARA A, RADEV D. A system for summarizing scientific topics starting from keywords[C]//Proceedings of the 51st annual meeting of the Association for Computational Linguistics. Sofia: Association for Computational Linguistics, 2013: 572 – 577.
[26] KHALID A, KHANAF A, IMRANB M, et al. Reference terms identification of cited articles as topics from citation contexts[J]. Computers and electrical engineering, 2018, 20(3): 1 – 12.
[27] GARZONE M, MERCER R E. Towards an automated citation classifier[C]//Advances in artificial intelligence. Berlin: Lecturenotes in computer science, 2000: 337 – 346.
[28] RADOULOV R. Exploring automatic citation classification[D]. Waterloo: University of Waterloo, 2008.
[29] RADEV D R, MUTHUKRISHNAN P, QAZVINIAN V, et al. The ACL anthology network corpus[J]. Language resources andevaluation, 2013, 47(4): 1 – 26.
[30] BIKEL D M, MILLER S, SCHWARTZ R, et al. 1997. Nymble: a high-performance learning name-finder[C]// Proceedings of the fifth conference on applied natural language processing. Washington: Association for Computational Linguistics, 1997: 194 – 201.
[31] LAFFERTY J D, MCCALLUM A, PEREIRA F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data[C]// Eighteenth international conference on machine learning. Williamsport: Morgan Kaufmann Publishers Inc, 2001: 282 – 289.
[32] MA X, HOVY E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF[C]//Proceedings of the 54th annual meeting of the association for computational linguistics. Berlin: Association for Computational Linguistics, 2016: 1064 – 1074.
[33] DEVLIN J, CHANG M W, LEEK, etal. BERT: Pre-training of deep bidirectional transformers for language understanding[EB/OL]. [2018 – 12 – 13]. https://arxiv.org/abs/1810.04805.
[34] PETERS M E, AMMAR W, BHAGAVATULA C, et al. Semi-supervised sequence tagging with bidirectional language models [C]// Proceedings of the 55th annual meeting of the Association for Computational Linguistics. Vancouver: Association for Computational Linguistics, 2017: 1756 – 1765.
[35] YANGZ L, SALAKHUTDINOV R, COHEN W W. Transfer learning for sequence tagging with hierarchical recurrent networks[C]// Proceedings of the 5th international conference on learning representations. Toulon: The DBLP Computer Science Bibliography, 2017.

作者贡献说明:

马娜:负责相关文献整理阅读和文章撰写;
张智雄:负责文章脉络把握和具体写作思路指导;
于改红:参与论文修改。

马娜, 张智雄, 于改红. 科技论文引用对象研究综述[J]. 图书情报工作,2019,63(23):139 – 145.

A Review of Citation Object Research

Ma Na[1,2] Zhang Zhixiong[1,2,3] Yu Gaihong[1,2]

[1] National Science Library, Chinese Academy of Sciences, Beijing 100190

[2] School of Economic and Management, University of Chinese Academy of Sciences, Beijing 100190

[3] Wuhan Library, Chinese Academy of Sciences, Wuhan 430071

Abstract: [**Purpose/significance**] In order to improve the effect of content-based citation analysis, this paper summarizes the related research on citation object at home and abroad, and provides reference for citation content analysts. [**Method/process**] By investigating the related research on citation object at home and abroad, this paper reviewed the definitions of the citation object, classification systems, application fields and automatic identification, summarized the current research on the citation object and put forward the future development direction. [**Result/conclusion**] The citation object evaluates the contribution and utilization value of academic research from the semantic level, which adds an important dimension to the citation analysis method. The research on citation object needs to be deepened in three directions: theoretically, to strengthen the research and analysis of multi-dimensional citation object features; technically, to explore the automatic identification methods based on large-scale corpus; application, to try to provide scientific research evaluation services based on citation objects.

Keywords: citation object citation object identify citation context analysis research evaluation

《图书情报工作》首届青年编委会成立暨座谈会在武汉举办

2019 年 12 月 6 日,在 2019 年图书情报与档案管理研究生教育论坛与青年学者论坛召开期间,在武汉大学信息管理学院的大力支持下,《图书情报工作》首届青年编辑委员会成立暨座谈会议在武汉大学信息管理学院举办。

座谈会首先由青年编委会主任、武汉大学信息管理学院副院长、教育部青年长江学者陆伟教授主持并发言,他认为青年编委会的成立不仅体现了《图书情报工作》杂志社开拓进取、勇于创新的精神,更重要的是体现了对青年学者的重视、关心和支持,青年人才的成长离不开期刊,无论是发表论文还是参与审稿都是个人学术成长道路中的重要步骤,并借此机会对《图书情报工作》杂志社表示感谢。《图书情报工作》杂志社社长兼主编、中国科学院大学图书情报与档案管理系主任初景利教授介绍了青年编委会的成立背景和章程、青年编委的遴选原则、权利与职责。首届青年编委会共邀请来自全国不同高校的 23 位青年才俊,旨在充分发挥他们的学术研究能力和创新能力,为期刊发展注入新的活力,同时也为青年学者提供发展和成长的平台。随后初主编向到会的 17 位青年编委颁发了聘书。接下来,在编委会副主任、华中师范大学信息管理学院副院长曹高辉副教授主持下,与会编委围绕《图书情报工作》改名设想、期刊定位、图情档学科的未来发展方向、学科建设与人才培养、期刊需要关注的问题、办刊举措等积极建言献策,展开了热烈研讨。

青年学者往往拥有良好的教育背景、旺盛的学术精力、敏锐的学术眼光,他们代表着图情学研究的未来,在学科和事业建设中被寄予重任和厚望。本次《图书情报工作》青年编委会的成立恰逢其时,期望编委们与和《图书情报工作》一起努力,为提高期刊质量和学术影响力,为推动业界学界的学术研究和图情档事业的发展,做出应有的贡献。

《图书情报工作》首届青年编委会名单